CrossMark

Circumcentering the Douglas–Rachford method

Roger Behling[1] · **José Yunier Bello Cruz**[2] ·
Luiz-Rafael Santos[1]

Abstract We introduce and study a geometric modification of the Douglas–Rachford method called the Circumcentered–Douglas–Rachford method. This method iterates by taking the intersection of bisectors of reflection steps for solving certain classes of feasibility problems. The convergence analysis is established for best approximation problems involving two (affine) subspaces and both our theoretical and numerical results compare favorably to the original Douglas–Rachford method. Under suitable conditions, it is shown that the linear rate of convergence of the Circumcentered–Douglas–Rachford method is at least the cosine of the Friedrichs angle between the (affine) subspaces, which is known to be the sharp rate for the Douglas–Rachford method. We also present a preliminary discussion on the Circumcentered–Douglas–Rachford method applied to the many set case and to examples featuring non-affine convex sets.

✉ Roger Behling
roger.behling@ufsc.br

José Yunier Bello Cruz
yunierbello@niu.edu

Luiz-Rafael Santos
l.r.santos@ufsc.br

[1] Department of Exact Sciences, Federal University of Santa Catarina, Blumenau, SC 89036-256, Brazil

[2] Department of Mathematical Sciences, Northern Illinois University, Watson Hall 366, DeKalb, IL 60115-2828, USA

Springer

Mathematics Subject Classification (2010) Primary 49M27 · 65K05 · 65B99;
Secondary 90C25

1 Introduction

Projection and reflection schemes are celebrated tools for finding a point in the
intersection of finitely many sets [15], a basic problem in the natural sciences and
engineering (see, e.g., [6] and [17]). Probably, the Douglas–Rachford method (DRM)
is one of the most famous and well-studied of these schemes (see, e.g., [4, 5, 8, 27,
30]). Also known as averaged alternating reflections method, it was introduced in
[22] and has recently gained much popularity, in part thanks to its good behavior in
non-convex settings (see, e.g., [1–3, 9, 11, 13, 24, 25, 28]).

This paper has a two-fold aim: (i) *improving* the original DRM by means of a
simple geometric modification, and (ii) meeting the demand of more satisfactory
schemes for the many set case.

Regarding (i), the proposed scheme is *greedy* by means of distance to the solution
set, that is, our iterate is the best possible point relying on successive reflections onto
two subspaces. In particular, we get an improvement towards the solution set with
respect to DRM iteration, which arises from the average of two successive reflec-
tions. Also, we get a convergence rate at least as good as DRM's with a comparable
computational effort per iteration, however with numerical results fairly favorable.
The aim (ii) will be treated in the last section.

To consider the problem, let $\langle \cdot, \cdot \rangle$ denote the scalar product in $\mathbb{R}^n$, $\| \cdot \|$ is the
induced norm, that is, the euclidean vector or matrix norm, and P_S denotes the
orthogonal projection onto a nonempty closed and convex set $S \subset \mathbb{R}^n$. The reflec-
tion onto S is given by means of P_S, namely, $R_S(x) = (2P_S - \mathrm{Id})(x)$, where Id
stands for the identity operator. Note that $P_S(x)$ is simply the midpoint of the segment
$[x, R_S(x)]$.

Our results are established for the following fundamental feasibility problem.
Given two *subspaces* U, V of $\mathbb{R}^n$ and any point $x \in \mathbb{R}^n$, we are interested in solving
the *best approximation problem* [20] of finding the closest point to x in $U \cap V$, i.e.,

$$\text{Find } \bar{x} \in U \cap V \text{ such that } \|\bar{x} - x\| = \min_{w \in U \cap V} \|w - x\|. \tag{1}$$

For subspaces, there exists a nice characterization of the best approximation problem
above as $\bar{x} = P_{U \cap V}(x)$ if, and only if, $x - \bar{x} \in (U \cap V)^\perp$, i.e.,

$$\langle x - \bar{x}, w \rangle = 0 \qquad \forall w \in U \cap V.$$

The classical DRM iteration at a point $x \in \mathbb{R}^n$ yields a new iterate $T(x) \in \mathbb{R}^n$
given by the midpoint between x and $R_V R_U(x)$. That is, the Douglas–Rachford (DR)
operator $T : \mathbb{R}^n \to \mathbb{R}^n$ designed to solve (1) reads as follows

$$T = T_{U,V} := \frac{\mathrm{Id} + R_V R_U}{2}.$$

It is known that, under suitable assumptions, DRM generates a sequence $\{T^k(x)\}_{k \in \mathbb{N}}$
converging to the solution of the best approximation problem (1) (see [4]).

Let us now introduce our scheme, called the Circumcentered–Douglas–Rachford method (C–DRM): from a point $x \in \mathbb{R}^n$, the next iterate is the circumcenter of the triangle of vertexes x, $y := R_U(x)$ and $z := R_V R_U(x)$, denoted by

$$C_T(x) := \text{circumcenter}\{x, y, z\}. \tag{2}$$

By circumcenter we mean that $C_T(x)$ is equidistant to x, y, and z and lies on the affine space defined by these vertexes. For all $x \in \mathbb{R}^n$, $C_T(x)$ exists, is unique and elementary computable. Existence and uniqueness are obvious if the cardinality of the set $\{x, y, z\}$ is either 1 or 2. In fact, if it happens that $x = y = z$, we have $C_T(x) = x \in U \cap V$ already—the converse is also true, that is, if $C_T(x) = x$, then $x = y = z$. This means that the set of fixed points Fix C_T of the (nonlinear) operator C_T is equal to $U \cap V$. If the cardinality $\{x, y, z\}$ is 2 then $C_T(x)$ is the midpoint between the two distinct points. If x, y, and z are distinct both existence and uniqueness follow from elementary geometry. Thus, one would only have to worry about having a situation in which x, y, and z are simultaneously distinct and collinear. This cannot happen since reflections onto subspaces are norm preserving. More than that, will further see that the distance of x, y, and z to $U \cap V$ is exactly the same. Thus, the equidistance we are asking for in (2) turns out to be a necessary condition for a solution of (1).

Note that C–DRM has indeed a two-dimensional search flavor: we will show that $C_T(x)$ is the closest point to $U \cap V$ belonging to aff$\{x, y, z\}$, the affine space defined by x, y, and z, whose dimension is equal to 2, if x, y, and z, are distinct. This property, together with the fact that the DR point $T(x) \in$ aff$\{x, y, z\}$, is the key to proving a better performance of C–DRM over DRM. An immediate consequence of this nice interpretation is that C–DRM will solve problems in $\mathbb{R}^2$ in at most two steps. Figure 1 serves as an intuition guide and illustrates our idea.

This paper is organized as follows. In Section 2, we derive the convergence analysis proving that C–DRM has a rate of convergence for solving problem (1) at least as good as DRM's. Section 3 presents numerical experiments for subspaces along with some non-affine examples in $\mathbb{R}^2$. Final remarks as how one can adapt C–DRM for the many set case and other comments on future work are presented in Section 4.

2 Convergence analysis of the Circumcentered–Douglas–Rachford method

In this section, we present the theoretical advantages of using C–DRM over the classical DRM iteration for solving (1).

In order to simplify the presentation we recall and fix some notation. For a point $x \in \mathbb{R}^n$, we denote for now on $y := R_U(x)$ and $z := R_V R_U(x)$. Recall that, at the point $x \in \mathbb{R}^n$, we define the C–DRM iteration as

$$C_T(x) := \text{circumcenter}\{x, y, z\}$$

and the DR iteration (at the same point) is given by

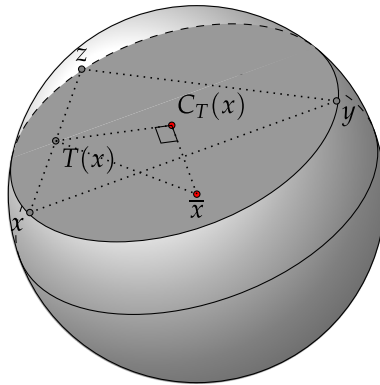$$T(x) := \frac{\text{Id} + R_V R_U}{2}(x) = \frac{x+z}{2}.$$

Fig. 1 Geometrical interpretation for circumcentering DRM

In the following, we present some elementary facts [20, Theorem 5.8] needed throughout the text.

Proposition 1 *Let S be a given subspace and $x \in \mathbb{R}^n$ arbitrary but fixed. Then, for all $s \in S$ we have:*

(i) $\langle x - P_S(x), s \rangle = 0$;
(ii) $\|x - P_S(x)\|^2 = \|x - s\|^2 - \|s - P_S(x)\|^2$;
(iii) $\|x - s\| = \|R_S(x) - s\|$;
(iv) *The projection and reflection mappings P_S and R_S are linear.*

Our first lemma states that the projection of y and z onto $U \cap V$ coincide with $P_{U \cap V}(x)$ and that the distances of x, y, and z to $U \cap V$ are the same. In addition to that, we have that the projection of the DR point $T(x)$ onto $U \cap V$ is as well given by $P_{U \cap V}(x)$.

Lemma 1 *Let $x \in \mathbb{R}^n$. Then,*

$$P_{U \cap V}(x) = P_{U \cap V}(y) = P_{U \cap V}(z)$$

and

$$\text{dist}(x, U \cap V) = \text{dist}(y, U \cap V) = \text{dist}(z, U \cap V).$$

Moreover, $P_{U \cap V}(x) = P_{U \cap V}(T(x))$ and

$$P_{U \cap V}(T^k(x)) := P_{U \cap V}(\underbrace{T(\cdots T(T(x)) \cdots)}_{k}).$$

Proof We consider a bar to denote the projection onto the subspace $U \cap V$, i.e., $\bar{x} := P_{U \cap V}(x)$, $\bar{y} := P_{U \cap V}(y)$, $\bar{z} := P_{U \cap V}(z)$, etc. By using Proposition 1(iii) for

the reflection onto U, we get $\|x - w\| = \|y - w\|$ for all $w \in U \cap V$. In particular, $\|x - \bar{y}\| = \|y - \bar{y}\|$ and $\|y - \bar{x}\| = \|x - \bar{x}\|$ since $\bar{x}, \bar{y} \in U \cap V$. Therefore,

$$\|x - \bar{x}\| \leq \|x - \bar{y}\| = \|y - \bar{y}\| \leq \|y - \bar{x}\| = \|x - \bar{x}\|,$$

which implies

$$\|x - \bar{x}\| = \|x - \bar{y}\| = \|y - \bar{x}\| = \|y - \bar{y}\|,$$

and, of course, $\text{dist}(x, U \cap V) = \text{dist}(y, U \cap V)$ and $P_{U \cap V}(x) = P_{U \cap V}(y)$ follow.

Now, the statements $\text{dist}(y, U \cap V) = \text{dist}(z, U \cap V)$ and $P_{U \cap V}(y) = P_{U \cap V}(z)$ can be derived by repeating the same argument with y and z, where Proposition 1(iii) is then considered for the reflection onto V. As the projection onto subspaces is linear (Proposition 1(iv)) and $T(x) := \frac{x+z}{2}$, we have

$$P_{U \cap V}(T(x)) = P_{U \cap V}\left(\frac{x+z}{2}\right) = \frac{P_{U \cap V}(x)}{2} + \frac{P_{U \cap V}(z)}{2} = \frac{\bar{x}}{2} + \frac{\bar{x}}{2} = \bar{x}.$$

The rest of the proof follows inductively.

Indeed, we proved $P_{U \cap V}(s) = P_{U \cap V}(T(s))$ for all $s \in \mathbb{R}^n$. Then, by setting $s := T^{k-1}(x)$, we get

$$P_{U \cap V}(T^{k-1}(x)) = P_{U \cap V}(T(T^{k-1}(x))) = P_{U \cap V}(T^k(x)),$$

proving the lemma.   □

We proceed by characterizing $C_T(x)$ as the projection of any point $w \in U \cap V$ onto the affine subspace defined by x, y, and z denoted by $\text{aff}\{x, y, z\}$.

Lemma 2 *Let $x \in \mathbb{R}^n$ and $W^x := \text{aff}\{x, y, z\}$. Then,*

$$P_{W^x}(w) = C_T(x),$$

for all $w \in U \cap V$. In particular, $P_{W^x}(P_{U \cap V}(x)) = C_T(x)$.

Proof Let $w \in U \cap V$ be given and set $p := P_{W^x}(w)$. Recall that $C_T(x)$ is defined by being the only point in W^x equidistant to x, y, and z. So, to prove the lemma, we just need to show that p is equidistant to x, y and z. By Proposition 1(ii), we have

$$\begin{aligned}
\|x - p\|^2 &= \|x - w\|^2 - \|w - p\|^2, \\
\|y - p\|^2 &= \|y - w\|^2 - \|w - p\|^2, \\
\|z - p\|^2 &= \|z - w\|^2 - \|w - p\|^2.
\end{aligned}$$

Proposition 1(iii) and Lemma 1 provided $\|x - w\| = \|y - w\| = \|z - w\|$. Hence, the above equalities imply $\|x - p\| = \|y - p\| = \|z - p\|$, which proves the result.   □

We have just seen that $C_T(x)$ is the closest point in $W^x = \text{aff}\{x, y, z\}$ to $U \cap V$. In particular, the circumcenter $C_T(x)$ lies at least as close to $U \cap V$ as the DR point $T(x)$.

The next result shows that compositions of $C_T(\cdot)$ do not change the projection onto $U \cap V$, that is, for any $x \in \mathbb{R}^n$ and $k \in \mathbb{N}$ we have

$$P_{U\cap V}(C_T^k(x)) := P_{U\cap V}(\underbrace{C_T(\cdots C_T(C_T(x))\cdots))}_{k} = P_{U\cap V}(x).$$

This is a usual feature of algorithms designed to solve best approximation problems.

Lemma 3 *Let $x \in \mathbb{R}^n$ and $k \in \mathbb{N}$. Then,*

$$P_{U\cap V}(C_T^k(x)) = P_{U\cap V}(C_T(x)) = P_{U\cap V}(x).$$

Proof Note that by proving the second equality, the first follows easily by induction on k, likewise the one in the proof of Lemma 1. Therefore, let us prove that $P_{U\cap V}(C_T(x)) = P_{U\cap V}(x)$. Consider again the abbreviation $\bar{x} := P_{U\cap V}(x)$ and $\bar{x}_c := P_{U\cap V}(C_T(x))$. From Lemma 2, we have that $P_{W^x}(\bar{x}) = C_T(x)$ and also $P_{W^x}(\bar{x}_c) = C_T(x)$. Thus, by Pythagoras it follows that

$$\|x - \bar{x}\|^2 = \|x - C_T(x)\|^2 + \|\bar{x} - C_T(x)\|^2, \tag{3}$$

$$\|x - \bar{x}_c\|^2 = \|x - C_T(x)\|^2 + \|\bar{x}_c - C_T(x)\|^2. \tag{4}$$

Now, using again Pythagoras, for the triangles with vertexes $x, \bar{x}, \bar{x}_c$ and $C_T(x), \bar{x}_c, \bar{x}$, we get

$$\|x - \bar{x}_c\|^2 = \|\bar{x} - \bar{x}_c\|^2 + \|x - \bar{x}\|^2 \tag{5}$$

and

$$\|C_T(x) - \bar{x}\|^2 = \|\bar{x} - \bar{x}_c\|^2 + \|C_T(x) - \bar{x}_c\|^2. \tag{6}$$

Then, we obtain

$$\|\bar{x} - \bar{x}_c\|^2 = \|x - \bar{x}_c\|^2 - \|x - \bar{x}\|^2 = \|\bar{x}_c - C_T(x)\|^2 - \|\bar{x} - C_T(x)\|^2 = -\|\bar{x} - \bar{x}_c\|^2,$$

where the first equality follows from (5), the second from subtracting (3) and (4) and the third follows from (6). Thus, $\|\bar{x} - \bar{x}_c\| = 0$, or equivalently, $\bar{x} = \bar{x}_c$. □

Note that Lemma 3 is related to *Fejér monotonicity* (see [7, Proposition 5.9 (i)]).

We will now measure the improvement of $C_T(x)$ towards $U \cap V$ by means of x and $T(x)$.

Lemma 4 *For each $x \in \mathbb{R}^n$, we have*

$$\text{dist}(C_T(x), U \cap V)^2 = \text{dist}(T(x), U \cap V)^2 - \|T(x) - C_T(x)\|^2. \tag{7}$$

In particular,

$$\text{dist}(C_T(x), U\cap V) = \|C_T(x) - P_{U\cap V}(x)\| \leq \|T(x) - P_{U\cap V}(x)\| = \text{dist}(T(x), U\cap V).$$

Proof Lemma 2 says in particular that $P_{W^x}(P_{U\cap V}(x)) = C_T(x)$, which is equivalent to saying that $\langle s - C_T(x), P_{U\cap V}(x) - C_T(x) \rangle = 0$ for all s on the affine subspace W^x. Now, taking $T(x)$ for the role of s and using Pythagoras, we get (7), since $P_{U\cap V}(C_T(x)) = P_{U\cap V}(T(x)) = P_{U\cap V}(x)$ due to Lemmas 1 and 3. The rest of the result is direct consequence of Lemmas 3 and 1. □

The linear rate of convergence we are going to derive for C–DRM is given by the cosine of the Friedrichs angle between U and V, defined below.

Definition 1 The *cosine of the Friedrichs angle* θ_F between U and V is given by

$$c_F(U, V) := \sup \left\{ \langle u, v \rangle \mid u \in U \cap (U \cap V)^{\perp}, \ v \in V \cap (U \cap V)^{\perp}, \ \|u\| \leq 1, \ \|v\| \leq 1 \right\}.$$

If context permits, we use just c_F instead of $c_F(U, V)$.

Next, we state some fundamental properties of the Friedrichs angle.

Proposition 2 *Let $U, V \subset \mathbb{R}^n$ be subspaces, then:*

(i) $0 \leq c_F(U, V) = c_F(V, U) = c_F(U^{\perp}, V^{\perp}) < 1$.
(ii) $c_F = \|P_V P_U - P_{U \cap V}\| = \|P_{V^{\perp}} P_{U^{\perp}} - P_{U^{\perp} \cap V^{\perp}}\|$.

Proof (i) See [19, Theorems 13 and 16]; (ii): See [20, Lemma 9.5(7)]. □

The following result is an elementary fact in Linear Algebra and will be used in sequel.

Proposition 3 *Let a subspace $S \subset \mathbb{R}^n$ be given. If $x, p \in \mathbb{R}^n$ are such that their midpoint*

$$s := \frac{x + p}{2} \in S,$$

then $\operatorname{dist}(x, S) = \operatorname{dist}(p, S)$.

Proof Let $s := \frac{x+p}{2} \in S$, $\hat{x} := P_S(x)$ and $\hat{p} := P_S(p)$. Set $\tilde{p} := \hat{x} + 2(s - \hat{x}) \in S$ and note that $\tilde{p}$ is defined in such a way that the triangle with vertexes x, $\hat{x}$, and s is congruent to the triangle formed by p, $\tilde{p}$, and s. In particular, $\|\tilde{p} - p\| = \|x - \hat{x}\|$. Therefore,

$$\|\hat{p} - p\| \leq \|\tilde{p} - p\| = \|x - \hat{x}\|. \tag{8}$$

An analogous construction can be considered for the triangle with vertexes p, $\hat{p}$, and s and the one with vertexes x, $\tilde{x}$, and s, where $\tilde{x} := \hat{p} + 2(s - \hat{p}) \in S$, yielding $\|\hat{x} - x\| \leq \|\tilde{x} - x\| = \|p - \hat{p}\|$. This, combined with (8), proves the proposition. □

The next lemma organizes and summarizes known properties of sequences generated by DRM, some of which will be important in the proof of our main theorem. It is worth noting that items (i) and (vi) are novel to our knowledge.

Lemma 5 *Let $x \in \mathbb{R}^n$ be given. Then, the following assertions for DRM hold:*

(i) $\operatorname{dist}(T^k(x), U + V) = \operatorname{dist}(x, U + V)$ *for all $k \in \mathbb{N}$, where $U + V = \operatorname{span}(U \cup V)$;*
(ii)   *The set* $\operatorname{Fix} T := \left\{ x \in \mathbb{R}^n \mid T(x) = x \right\}$ *is given by* $\operatorname{Fix} T = (U \cap V) \oplus (U^{\perp} \cap V^{\perp})$;

(iii) *The DRM sequence $\{T^k(x)\}_{k\in\mathbb{N}}$ converges to $P_{\text{Fix}\,T}(x)$ and for all $k \in \mathbb{N}$,*

$$\|T^k(x) - P_{\text{Fix}\,T}(x)\| \leq c_F^k \|x - P_{\text{Fix}\,T}(x)\|;$$

(iv) *For all $k \in \mathbb{N}$ we have $P_{U\cap V}(T^k(x)) = P_{U\cap V}(x)$ and $P_{\text{Fix}\,T}(T^k(x)) = P_{\text{Fix}\,T}(x)$;*

(v) *$P_{U\cap V}(x) = P_{\text{Fix}\,T}(x)$ if, and only if, $x \in U + V$;*

(vi) *The DRM sequence $\{T^k(x)\}_{k\in\mathbb{N}}$ converges to $P_{U\cap V}(x)$ if, and only if,*

$$x \in U + V;$$

(vii) *The shadow sequences*

$$\{P_U(T^k(x))\}_{k\in\mathbb{N}} \text{ and } \{P_V(T^k(x))\}_{k\in\mathbb{N}}$$

converge to $P_{U\cap V}(x)$.

Proof For the sake of notation set $S := U + V$ and remind that $y := R_U(x)$ and $z := R_V(y)$. We have $\frac{1}{2}(x+y) = P_U(x) \in S$ and $\frac{1}{2}(y+z) = P_V(y) \in S$. Employing Proposition 3 yields $\text{dist}(x, S) = \text{dist}(y, S) = \text{dist}(z, S)$. Also, $\frac{1}{2}(T(x) + y) = \frac{1}{2}\left(\frac{x+z}{2} + y\right) = \frac{1}{2}(P_U(x) + P_V(y)) \in S$. Using again Proposition 3, we conclude that $\text{dist}(T(x), S) = \text{dist}(y, S)$. Hence, $\text{dist}(T(x), S) = \text{dist}(x, S)$ for all $x \in \mathbb{R}^n$. A simple induction procedure gives us $\text{dist}(T^k(x), U + V) = \text{dist}(x, U + V)$ for all $k \in \mathbb{N}$, proving (i).

The proofs of items (ii), (iii), (iv), and (vii) are in [4].

It is straightforward to check that $S^\perp = U^\perp \cap V^\perp$. Therefore, Fix T specialized to S reduces to $U \cap V$ and (v) follows. (vi) is a combination of (ii) and (v). □

We are now in the position to present our main convergence result, which states that the best approximation problem (1) can be solved for any point $x \in \mathbb{R}^n$ with usage of C–DRM.

Theorem 1 *Let $x \in \mathbb{R}^n$ be given. Then, the three C–DRM sequences*

$$\left\{C_T^k(P_U(x))\right\}_{k\in\mathbb{N}}, \left\{C_T^k(P_V(x))\right\}_{k\in\mathbb{N}} \text{ and } \left\{C_T^k(P_{U+V}(x))\right\}_{k\in\mathbb{N}}$$

converge linearly to $P_{U\cap V}(x)$. Moreover, their rate of convergence is at least the cosine of the Friedrichs angle $c_F \in [0, 1)$.

Proof Obviously, $u, v, s \in U + V$, with $u := P_U(x)$, $v := P_V(x)$ and $s := P_{U+V}(x)$. Let us first prove that $\bar{u} = \bar{v} = \bar{x} := P_{U\cap V}(x)$, where $\bar{u} := P_{U\cap V}(u)$, $\bar{v} := P_{U\cap V}(v)$ and $\bar{s} := P_{U\cap V}(s)$. The definition of $\bar{u}, \bar{x}$ allow us to employ Pythagoras and conclude that

$$\|u - x\|^2 + \|u - \bar{x}\|^2 = \|x - \bar{x}\|^2 \leq \|x - \bar{u}\|^2 = \|u - x\|^2 + \|u - \bar{u}\|^2,$$

which provides $\|u - \bar{x}\| = \|u - \bar{u}\|$. Thus $\bar{u} = \bar{x}$. We get $\bar{v} = \bar{s} = \bar{x}$ analogously, and by item (v) of Lemma 5 we further have $\bar{u} = P_{\mathrm{Fix}\,T}(u) = \bar{v} = P_{\mathrm{Fix}\,T}(v) = \bar{s} = P_{\mathrm{Fix}\,T}(s) = \bar{x}$. Hence, it holds that

$$
\begin{aligned}
\|C_T(u) - \bar{x}\| = \|C_T(u) - \bar{u}\| &\leq \|T(u) - \bar{u}\| = \|T(u) - P_{\mathrm{Fix}\,T}(u)\| \\
&\leq c_F \|u - P_{\mathrm{Fix}\,T}(u)\| = c_F \|u - \bar{x}\|,
\end{aligned}
$$

where the first inequality is by Lemma 4 and the second one by item (iii) of Lemma 5. Recall that Lemma 4 stated that $P_{U \cap V}(C_T^k(P_U(x))) = P_{U \cap V}(P_U(x))$ for all $k \in \mathbb{N}$. So, $P_{U \cap V}(C_T^k(u)) = P_{U \cap V}(u) = \bar{x}$ for all $k \in \mathbb{N}$. Consider now the induction hypothesis $\|C_T^{k-1}(u) - \bar{x}\| \leq c_F^{k-1}\|u - \bar{x}\|$ for some $k - 1$ — the case $k - 1 = 1$ was seen above — and note that

$$
\begin{aligned}
\|C_T^k(u) - \bar{x}\| &= \|C_T(C_T^{k-1}(u)) - P_{U \cap V}(C_T^{k-1}(u))\| \\
&\leq \|T(C_T^{k-1}(u)) - P_{U \cap V}(C_T^{k-1}(u))\| \\
&\leq c_F \|C_T^{k-1}(u) - P_{U \cap V}(C_T^{k-1}(u))\| \\
&= c_F \|C_T^{k-1}(u) - \bar{x}\| \leq c_F c_F^{k-1}\|u - \bar{x}\| = c_F^k \|u - \bar{x}\|,
\end{aligned}
$$

where the first inequality is due to Lemma 4. The second inequality above follows from Lemma 5 (iii) and (v), since $u \in U + V$, and the third is by the induction hypothesis. This proves the theorem for the sequence $\{C_T^k(P_U(x))\}$. The proof lines for the convergence of the sequences $\{C_T^k(P_V(x))\}$ and $\{C_T^k(P_{U+V}(x))\}$ with the linear rate c_F are analogous. $\qquad\square$

An immediate consequence is stated below.

Corollary 1 *Let $x \in U + V$ be given. Then, the C–DRM sequence $\{C_T^k(x)\}_{k \in \mathbb{N}}$ converges linearly to $P_{U \cap V}(x)$. Moreover, the rate of convergence is at least the cosine of the Friedrichs angle $c_F \in [0, 1)$.*

Although we believe that Theorem 1 holds for $\{C_T^k(x)\}$ itself, it is worth mentioning that considering the initial feasibility step $P_U(x)$ or $P_V(x)$ is totally reasonable, since we are assuming that the projection/reflection operators P_U, P_V, R_U, and R_V are available. In addition to that, these feasibility steps keep the whole C–DRM sequence in $U + V$, which can be very desirable. In this sense, let us look at two distinct lines $U, V \subset \mathbb{R}^3$ intersecting at the origin and assume that their Friedrichs angle is not ninety degrees, i.e., $c_F \in (0, 1)$. C–DRM finds the origin after one or two steps when starting in $U + V$, since $U + V$ is the plane containing the two lines. If the initial point is not in $U + V$ and no feasibility step is taken, C–DRM may generate an infinite sequence. Therefore, running C–DRM in $U + V$, a potentially smaller set than $\mathbb{R}^n$, sounds attractive from the numerical point of view.

The feasibility procedure employed in Theorem 1 provides convergence to best approximation solutions for the DRM without the need of considering shadow sequences. This is formally presented in the following.

Corollary 2 *Let $x \in \mathbb{R}^n$ be given. Then, the three DRM sequences*

$$\{T^k(P_U(x))\}_{k\in\mathbb{N}}, \ \{T^k(P_V(x))\}_{k\in\mathbb{N}} \ and \ \{T^k(P_{U+V}(x))\}_{k\in\mathbb{N}}$$

converge linearly to $P_{U\cap V}(x)$. Moreover, their rate of convergence is given by the cosine of the Friedrichs angle $c_F \in [0, 1)$.

Proof This result follows from the fact that $P_{U\cap V}(P_U(x)) = P_{U\cap V}(P_V(x)) = P_{U\cap V}(P_{U+V}(x)) = P_{U\cap V}(x)$, established within the proof of Theorem 1, combined with Lemma 5 (vi). □

It is known that c_F is the sharp rate for DRM [4]. This is not clear for C–DRM though and left as an open question in this work. One way of addressing this issue would be looking at the magnitude of improvement of C–DRM over DRM given by (7) in Lemma 4.

We finish this section by showing that Theorem 1 and Corollaries 1 and 2 are applicable to affine subspaces with nonempty intersection, where the concept of the cosine of the Friedrichs angle is suitably extended.

Corollary 3 *Let A and B be affine subspaces of $\mathbb{R}^n$ with nonempty intersection and $p \in A \cap B$ arbitrary but fixed. Then, Theorem 1and Corollaries 1 and 2 hold for A and B, with the rate c_F being the cosine of the Friedrichs angle between the subspaces $U_A := A - \{p\}$ and $V_B := B - \{p\}$.*

Proof Since $p \in A \cap B$, the translations $A - \{p\}$ and $B - \{p\}$ provide nonempty subspaces U_A and V_B. Now, the elementary translation properties of reflections $R_{U_A}(x) + p = R_A(x + p)$ and $R_{V_B}(x) + p = R_B(x + p)$ give us the translation formulas for the correspondent Douglas-Rachford and circumcentering operators:

$$T_{A,B}(x + p) := T_{U_A,V_B}(x) + p$$

and

$$C_{T_{A,B}}(x + p) := C_{T_{U_A,V_B}}(x) + p,$$

for all $x \in \mathbb{R}^n$. This suffices to prove the corollary when setting $U := U_A$ and $V := V_B$ in Theorem 1 and Corollaries 1 and 2, because the above formulas imply that $T_{A,B}^k(x+p) = T_{U_A,V_B}^k(x)+p$ and $C_{T_{A,B}}^k(x+p) = C_{T_{U_A,V_B}}^k(x)+p$, for all $k \in \mathbb{N}$. □

Simple manipulations let us conclude that $c_F = c_F(U_A, V_B)$, with U_A and V_B as above, does not depend on the particular choice of the point p in $A\cap B$. Therefore, c_F can be referred to as the cosine of the Friedrichs angle between the affine subspaces A and B with no ambiguity.

We finish this section remarking that the computation of circumcenters is possible in arbitrary inner product spaces—see (9). However, projecting/reflecting onto subspaces depends on their closedness. Now, in Hilbert spaces, it is well known that having c_F strictly smaller than 1 is equivalent to asking $U + V$ to be closed [19, Theorem 13]. This is an assumption that maintains the linear convergence in Hilbert spaces

for several projection/reflection methods and that would also enable us to extend our main results concerning C–DRM to an infinite dimensional setting.

The following section is on numerical experiments and begins by showing how one can compute $C_T(x)$ by means of elementary and cheap Linear Algebra operations.

3 Numerical experiments

In this section, we make use of numerical experiments to show, as a proof of concept, that the good theoretical proprieties of $C_T(x)$ are also working in practical problems.

First, for a given point $x \in \mathbb{R}^n$, we establish a procedure to find $C_T(x)$. We then discuss the stopping criteria used in our experiments and show the performance of C–DRM compared to DRM applied to problem (1). We also apply C–DRM and DRM to non-affine samples, which are not treated theoretically in the previous section. The experiments with these problems indicate as well a good behavior of C–DRM over DRM. All experiments were performed using `julia` [12] programming language and the pictures were generated using PGFPlots [23].

3.1 How to compute the circumcenter in $\mathbb{R}^n$

In order to compute $C_T(x)$, recall that $y := R_U(x)$ and $z := R_V(y) = R_V R_U(x)$. We have already mentioned that $C_T(x) = x$ if, and only if, the cardinality of aff$\{x, y, z\}$ is 1. If this cardinality is 2, $C_T(x)$ is the midpoint between the two distinct points.

Therefore, we will focus on the computation of $C_T(x)$ for the case where x, y and z are distinct. According to Lemma 2, $C_T(x)$ is still well defined, and it can be seen as the circumcenter of the triangle Δxyz formed by the points x, y, and z, as illustrated in Fig. 2. Also, $T(x)$ lies in aff$\{x, y, z\}$.
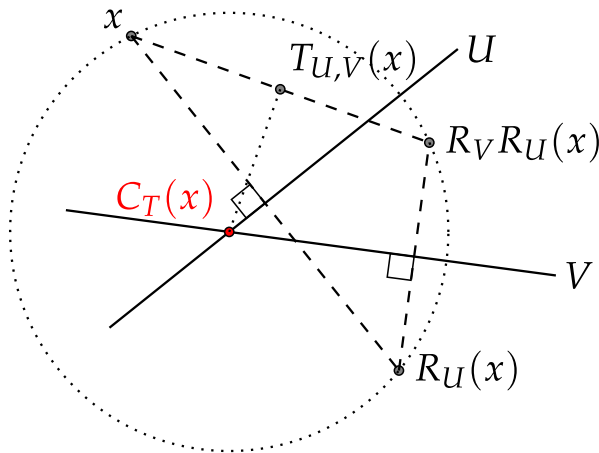

Fig. 2 Circumcenter on the affine subspace aff$\{x, y, z\}$

Let x be *the anchor* point of our framework and define $s_U := y - x$ and $s_V :$ $= z - x$, the vectors pointing from x to y and from x to z, respectively. Note that $\operatorname{aff}\{x, y, z\} = x + \operatorname{span}\{s_U, s_V\}$, and since $C_T(x) \in \operatorname{aff}\{x, y, z\}$, the dimension of the ambient space, namely n, is irrelevant to the geometry. The problem is *intrinsically* a two-dimensional one, regardless of what n is.

We want to find the vector $s \in \operatorname{aff}\{x, y, z\}$, whose projection onto each vector s_U and s_V has its endpoint at the midpoint of the line segment from x to y and x to z, that is,

$$P_{\operatorname{span}\{s_U\}}(s) = \frac{1}{2}s_U \quad \text{and} \quad P_{\operatorname{span}\{s_V\}}(s) = \frac{1}{2}s_V.$$

This requirement yields

$$\begin{cases} \langle s_U, s \rangle = \frac{1}{2}\|s_U\|^2, \\ \langle s_V, s \rangle = \frac{1}{2}\|s_V\|^2. \end{cases}$$

By writing $s = \alpha s_U + \beta s_V$, we get the 2×2 linear system with unique solution (α, β)

$$\begin{cases} \alpha\|s_U\|^2 + \beta \langle s_U, s_V \rangle = \frac{1}{2}\|s_U\|^2, \\ \alpha \langle s_U, s_V \rangle + \beta\|s_V\|^2 = \frac{1}{2}\|s_V\|^2. \end{cases} \tag{9}$$

Hence,

$$C_T(x) = x + \alpha s_U + \beta s_V.$$

Note that (9) enables us to calculate $C_T(x)$ in arbitrary inner product spaces.

3.2 The case of two subspaces

In our experiments, we randomly generate 100 instances with subspaces U and V in $\mathbb{R}^{200}$ such that $U \cap V \neq \{0\}$. Each instance is run for 20 initial random points. Based on Theorem 1 and Corollary 2, we monitor the C–DRM sequence $\{C_T^k(P_V(x))\}$ and the DRM sequence $\{T^k(P_V(x))\}$. Note that the DRM sequence will always converge to $U \cap V$, since $P_V(x) \in U + V$. For the C–DRM sequence, such hypothesis does not seem to be necessary, however for the sake of fairness, we choose to monitor this sequence in the same way. We incorporate the method of alternating projections (MAP) (see, e.g., [18]) in our experiments as well. MAP generates a sequence $\{(P_V P_U)^k(x)\}$ that lies automatically in $U + V$ for all $k \geq 1$.

Let $\{s^k\}$ be any of the three sequences that we monitor. We considered two tolerances

$$\varepsilon_1 := 10^{-3} \text{ and } \varepsilon_2 := 10^{-6},$$

and employed two stopping criteria:

– the *true error*, given by $\|s^k - \bar{x}\| < \varepsilon_i$; and
– the *gap distance*, given by $\|P_U(s^k) - P_V(s^k)\| < \varepsilon_i$, for $i = 1, 2$.

We performed tests with two tolerances in order to challenge C–DRM by asking for more precision. Also, one can view the *true error* as the best way to assure one is sufficiently close to $U \cap V$, and in our case $\bar{x} := P_{U \cap V}(x)$ is easily available. However, this is not the case in applications, thus, we also utilized the *gap distance*, which we consider a reasonable measure of infeasibility.

In order to represent the results of our numerical experiments, we used the Performance Profiles [21], a analytic tool that allows one to compare several different algorithms on a set of problems with respect to a performance measure or *cost*, which in our case is the number of iterations, providing the visualization and interpretation of the results of benchmark experiments. The rationale of choosing number of iterations as performance measure here is that in each of the sequences that are monitored, the majority of the computational cost involved is equivalent—two orthogonal projections onto the subspaces U and V per iteration have to be computed for each method. The graphics were generated using perprof-py [29].

The numerical experiments shown in Fig. 3 confirm the theoretical results obtained in this paper, since C–DRM has a much better performance than DRM. For $\varepsilon_2 = 10^{-6}$, C–DRM solves all instances and choices of initial points in less iterations than DRM, regardless the stopping criteria (See Fig. 3b and d). For $\varepsilon_1 = 10^{-3}$, using the true error criteria, according to Fig. 3a, a tiny part of the instances were solved
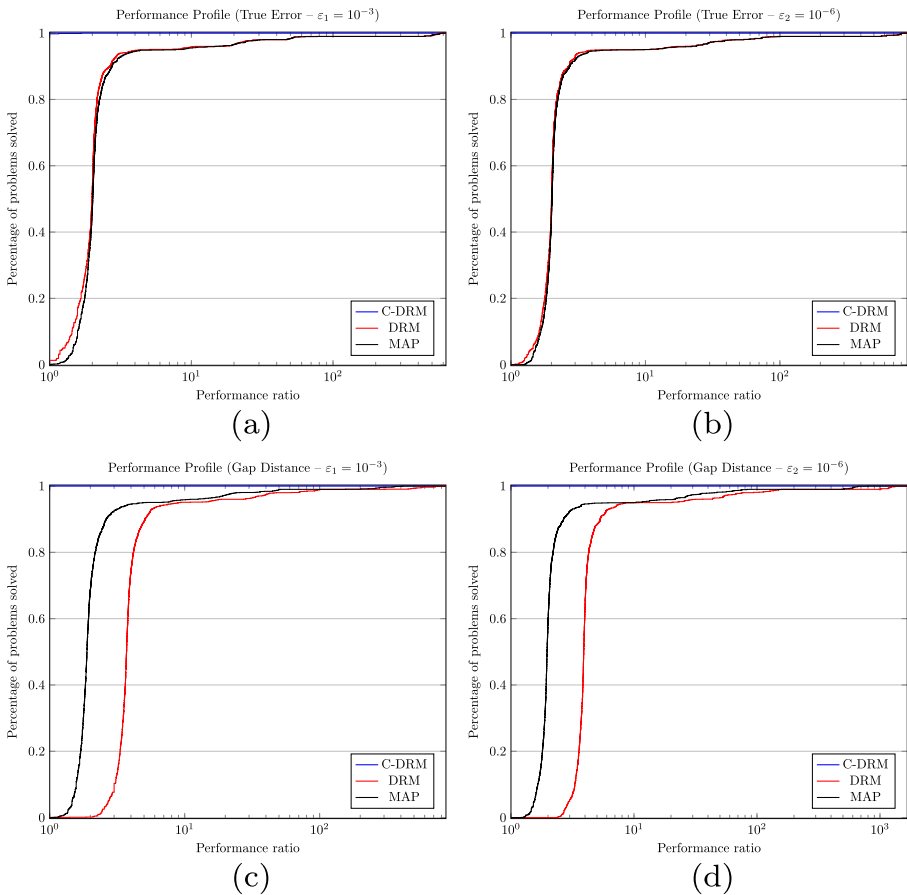

Fig. 3 Experiments with two subspaces

faster by DRM; however, this behavior was not reproduced with the gap distance criteria (see Fig. 3d). MAP has c_F^2 as asymptotic linear rate (see [26]) and it was beaten by C–DRM in all our instances. This gives rise to the interesting question of whether the rate c_F^2 can be theoretically achieved by C–DRM.

Figure 4 represents experiments in which the Friederichs angle between the two subspaces is smaller than 10^{-2} and the *true error* criterion is used. In this case, MAP and DRM are known to converge slowly. C–DRM, however, handled the small values of the Friederichs angle substantially better.

3.3 Some non-affine examples

Next, we present two simple classes of examples revealing the potential of the proposed modification when it is applied to the convex and the non-convex case. Here, we are using the *gap distance* with $\varepsilon_2 = 10^{-6}$.

Example 1 (Two balls in $\mathbb{R}^2$) We present three figures that depict numerical experiments showing the behavior of C–DRM over DRM concerning the problem of finding a point in the intersection of two convex balls in $\mathbb{R}^2$.

Note that Lemma 4 shows that $C_T(x)$ is closer to $U \cap V$ than $T(x)$ for problem (1). Unfortunately, this is not true in general for convex sets. Figure 5a, illustrates this for two balls. Here, x^0 is the starting point, $\bar{x} = P_{U \cap V}(x^0)$ is the only point in the intersection, $x_C^1 := C_T(x^0)$ and $x_{DR}^1 := T(x^0)$. Note however, that this does not mean that C–DRM will not work for the general case. Even though x_{DR}^1 is closer to $\bar{x}$ than x_C^1 is, C–DRM performed way less iterations (37) than DRM (971) to find the solution.

Below we present two examples, in Fig. 5b, c, where the two balls intersection is still compact but has infinitely many points. Depending on the starting point, we achieve different results, regarding the number of iterations that both C–DRM and DRM take to converge. Moreover, Fig. 5b shows that the accumulation point of the
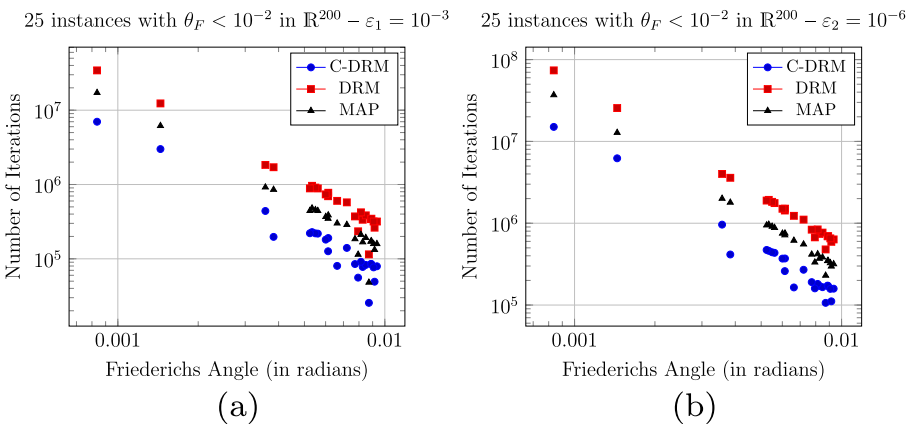

25 instances with $\theta_F < 10^{-2}$ in $\mathbb{R}^{200} - \varepsilon_1 = 10^{-3}$ 25 instances with $\theta_F < 10^{-2}$ in $\mathbb{R}^{200} - \varepsilon_2 = 10^{-6}$

(a)   (b)

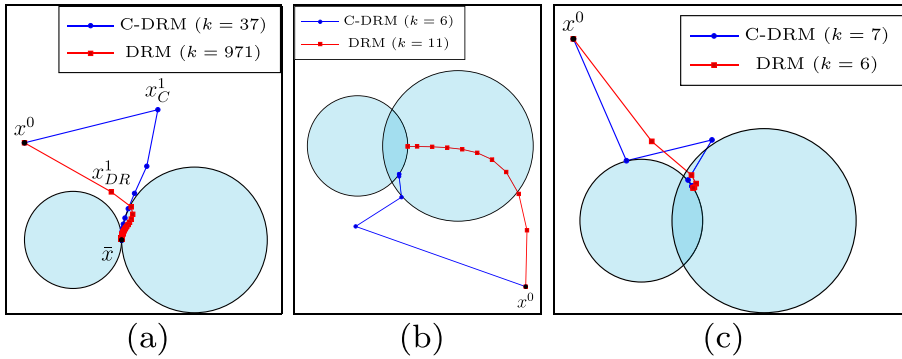**Fig. 4** Experiments with two subspaces having a small Friederichs angle

Fig. 5 Experiments with two balls in $\mathbb{R}^2$

sequences generated by C–DRM and DRM are not necessarily the same, with comparable iteration complexity though. In Fig. 5c, DRM converged after six iterations while C–DRM took seven iterations.

We performed extensive experiments featuring two convex balls in $\mathbb{R}^n$ with starting points being chosen randomly, and the results were very similar to the ones presented in the pictures above. These positive experiments show that it might be possible to use C–DRM to find a point in the intersection of non-affine convex sets. This should be sought in the future. Note however, that C–DRM need not be defined in the general convex setting. This may be the case when an iterate happens to reach the line passing through the centers of the balls in our examples. Therefore, a hybrid strategy is necessary in order to have C–DRM generating an infinite sequence.

Example 2 (A ball and a line and a circle and a line in $\mathbb{R}^2$)

Our examples show that C–DRM is likely to converge in less iterations than DRM. Also, as important as algorithmic complexity, is convergence to a solution itself.

In this regard, we underline that our pictures clearly display that C–DRM converges faster than DRM to a solution of problem (1). Moreover, DRM fails to find the unique common point of the ball and the line (only converging in the matter of shadow sequences) and to find the best approximation point in Fig. 6a, b, respectively.

Figure 6c represents a slightly non-convex example with a circle and a line, which was considered for DRM in [11]. In this experiment, both C–DRM and DRM converge and the latter was slower. This reveals an interesting and promising property of C–DRM, as well as DRM, when it is applied to non-convex problems. To end this discussion, it is worth mentioning that we performed extensive numerical tests for particular instances of non-convex problem in $\mathbb{R}^n$ and the results are very positive. This is a humble attempt in targeting the problem of finding a point in the intersection of an affine subspace with the s-space vectors defined by the generalized ℓ_0-ball (see problem (6) of [25]).
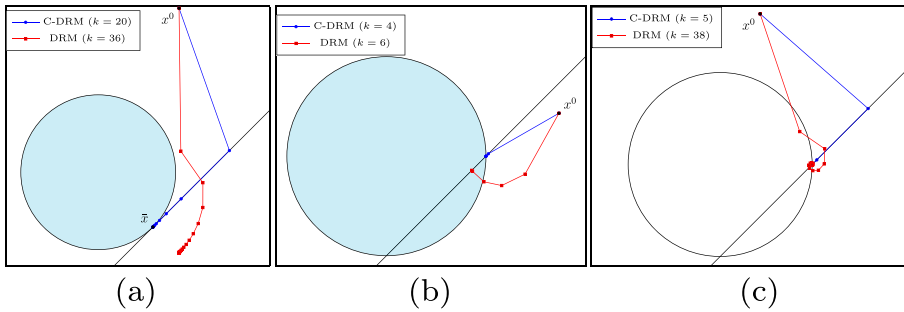
Fig. 6 Experiments with a ball and a line in $\mathbb{R}^2$

4 Conclusions and future work

We have introduced and derived a convergence analysis for the Circumcentered–Douglas–Rachford method (C–DRM) for best approximation problems featuring two (affine) subspaces $U, V \subset \mathbb{R}^n$. For any initial point, linear convergence to the best solution was shown and the rate of convergence of C–DRM was proven to be at least as good as DRM's. DRM is known to converge with the sharp rate $c_F \in [0, 1)$, the cosine of the Friedrichs angle between U and V (see, e.g., [4]). A question, left as open in this paper, is whether c_F is a sharp rate for C–DRM. In this regard, our numerical experiments show that *circumcentering* "speeds up" convergence of DRM for two subspaces as well as for most of our simple non-affine examples.

In view of future work, we end the paper with a brief discussion on the many set case and on how one can "circumcenter" other classical projection/reflection type methods.

The many set case Another relevant advantage of the circumcentering scheme C–DRM is that it can be extended to the following many set case.

Assume that $\{U_i\}_{i=1}^m \subset \mathbb{R}^n$ is a family of finitely many affine subspaces with nonempty intersection $\cap_{i=1}^m U_i$ and that we are interested in projecting onto $\cap_{i=1}^m U_i$ using knowledge provided by reflections onto each U_i.

For an arbitrary initial point $x \in \mathbb{R}^n$, we could consider a generalized C–DRM iteration by taking the circumcenter $C_T(x)$ of

$$\{x, R_{U_1}(x), R_{U_2}R_{U_1}(x), \ldots, R_{U_m}\cdots R_{U_2}R_{U_1}(x)\},$$

i.e., $C_T(x)$ is in aff$\{x, R_{U_1}(x), R_{U_2}R_{U_1}(x), \ldots, R_{U_m}\cdots R_{U_2}R_{U_1}(x)\}$ and

$$\mathrm{dist}(C_T(x), x) = \mathrm{dist}(C_T(x), R_{U_1}(x)) = \cdots = \mathrm{dist}(C_T(x), R_{U_m}\cdots R_{U_2}R_{U_1}(x)).$$

The fact that $C_T(x)$ is well defined and that it is precisely given by the projection of any point in $\cap_{i=1}^m U_i$ onto

$$W^x := \mathrm{aff}\{x, R_{U_1}(x), R_{U_2}R_{U_1}(x), \ldots, R_{U_m}\cdots R_{U_2}R_{U_1}(x)\}$$

can be derived similarly as in Lemma 2. We understand though, that the convergence analysis of C–DRM for the many set case might be substantially more challenging,

since we no longer can rely on the theory of DRM. Also, one should note that now, for large m, the computation of $C_T(x)$ may not be negligible. This computation consists of finding the intersection of m bisectors in W^x. Therefore, linear system (9) is now $m \times m$, and the calculation of $C_T(x)$ requires its resolution.

If, for a given problem, the computation of $C_T(x)$ mentioned above is simply too demanding, one could, e.g., consider pairwise circumcenters or even other ways of circumcentering. The important thing here is that we can enforce several strategies to help overcome the unsatisfactory extension of the classical Douglas–Rachford method for the many set case. It is known that for an example in $\mathbb{R}^2$ with three distinct lines intersecting at the original (see [1, Example 2.1]), the natural extension of the Douglas–Rachford method may fail to find a solution. On the other hand, any reasonable circumcentering scheme will solve this particular problem in at most two steps for any initial point. Circumcentering schemes may also be embedded in methods for the many set case, e.g., CADRA [10] and the Cyclic–Douglas–Rachford method [14].

Circumcentering other reflection/projection type methods For the case of U and V being affine subspaces, the reflected Cimmino method [16] considers a current point $x \in \mathbb{R}^n$ and takes the next iterate as the mean $\frac{1}{2}(R_U(x) + R_V(x))$. Circumcentering the Cimmino method is possible by setting the next iterate as circumcenter$\{x, R_U(x), R_V(x)\}$. Something similar can be done for the Method of Alternating Projections (MAP) (see, e.g., [18]). From a point $x \in \mathbb{R}^n$, MAP moves to $P_V P_U(x)$. In order to circumcenter MAP, one could take the circumcenter of x, $R_U(x)$ and $R_V P_U(x)$. The coherence of this approach lies on the fact that the mid points of the segments $[x, R_U(x)]$ and $[P_U(x), R_V P_U(x)]$ are $P_U(x) \in U$ and $P_V P_U(x) \in V$, respectively. Note that both the Circumcentered–Cimmino–Method and the Circumcentered–MAP solve problems with affine subspaces $U, V \subset \mathbb{R}^2$ in at most two steps. This happens because circumcentering can be seen as a 2-dimensional hyperplane search for the two set case.

Finally, we consider that investigating C–DRM for general convex feasibility problems [6] may be fruitful.

References

1. Aragón Artacho, F.J., Borwein, J.M., Tam, M.K.: Recent results on Douglas–Rachford methods for combinatorial optimization problems. J. Optim. Theory Appl. **163**(1), 1–30 (2014)
2. Aragón Artacho, F.J., Borwein, J.M., Tam, M.K.: Global behavior of the Douglas–Rachford method for a nonconvex feasibility problem. J. Glob. Optim. **65**(2), 309–327 (2016)
3. Aragón Artacho, F.J., Borwein, J.M.: Global convergence of a non-convex Douglas–Rachford iteration. J. Glob. Optim. **57**(3), 753–769 (2013)
4. Bauschke, H.H., Bello Cruz, J.Y., Nghia, T.T.A., Phan, H.M., Wang, X.: The rate of linear convergence of the Douglas–Rachford algorithm for subspaces is the cosine of the Friedrichs angle. J. Approx. Theory **185**, 63–79 (2014)

5. Bauschke, H.H., Bello Cruz, J.Y., Nghia, T.T.A., Phan, H.M., Wang, X.: Optimal rates of linear convergence of relaxed alternating projections and generalized Douglas-Rachford methods for two subspaces. Numer. Algor. **73**(1), 33–76 (2016)
6. Bauschke, H.H., Borwein, J.M.: On projection algorithms for solving convex feasibility problems. Siam Rev. **38**(3), 367–426 (2006)
7. Bauschke, H.H., Combettes, P.L.: Convex Analysis and Monotone Operator Theory in Hilbert Spaces, 1 edn. CMS Books in Mathematics. Springer International Publishing, Cham (2017)
8. Bauschke, H.H., Moursi, W.M.: On the Douglas–Rachford algorithm. Math. Program. **164**(1-2), 263–284 (2016)
9. Bauschke, H.H., Noll, D.: On the local convergence of the Douglas–Rachford algorithm. Arch. Math. **102**(6), 589–600 (2014)
10. Bauschke, H.H., Noll, D., Phan, H.M.: Linear and strong convergence of algorithms involving averaged nonexpansive operators. J. Math. Anal. Appl. **421**(1), 1–20 (2015)
11. Benoist, J.: The Douglas–Rachford algorithm for the case of the sphere and the line. J. Glob. Optim. **63**(2), 363–380 (2015)
12. Bezanson, J., Edelman, A., Karpinski, S., Shah, V.B.: Julia: A fresh approach to numerical computing. Siam Rev. **59**(1), 65–98 (2017)
13. Borwein, J.M., Sims, B.: The Douglas–Rachford algorithm in the absence of convexity. In: Fixed-Point Algorithms for Inverse Problems in Science and Engineering, pp. 93–109. Springer, New York (2011)
14. Borwein, J.M., Tam, M.K.: A cyclic Douglas–Rachford iteration scheme. J. Optim. Theory Appl. **160**(1), 1–29 (2014)
15. Censor, Y., Cegielski, A.: Projection methods: an annotated bibliography of books and reviews. Optimization **64**(11), 2343–2358 (2014)
16. Cimmino, G.: Calcolo approssimato per le soluzioni dei sistemi di equazioni lineari. Ric. Sci. **9**(II), 326–333 (1938)
17. Combettes, P.L.: The foundations of set theoretic estimation. Proc. IEEE **81**(2), 182–208 (1993)
18. Deutsch, F.R.: Rate of convergence of the method of alternating projections. In: Brosowski, B., Deutsch, F.R. (eds.) Parametric Optimization and Approximation: Conference Held at the Mathematisches Forschungsinstitut, Oberwolfach, October 16–22, 1983, pp. 96–107. Basel, Birkhäuser (1985)
19. Deutsch, F.R.: The angle between subspaces of a hilbert space. In: Approximation Theory, Wavelets and Applications, pp. 107–130. Springer, Dordrecht (1995)
20. Deutsch, F.R.: Best Approximation in Inner Product Spaces. CMS Books in Mathematics. Springer, New York (2001)
21. Dolan, E.D., Moré, J.J.: Benchmarking optimization software with performance profiles. Math. Program. **91**(2), 201–213 (2002)
22. Douglas, J., Rachford, H.H.: On the numerical solution of heat conduction problems in two and three space variables. Trans. Am. Math. Soc. **82**(2), 421–421 (1956)
23. Feuersänger, C.: PGFPlots Package, 1.13 edn (2016)
24. Hesse, R., Luke, D.R.: Nonconvex notions of regularity and convergence of fundamental algorithms for feasibility problems. SIAM J. Optim. **23**(4), 2397–2419 (2013)
25. Hesse, R., Luke, D.R., Neumann, P.: Alternating projections and Douglas-Rachford for sparse affine feasibility. IEEE Trans. Signal Process. **62**(18), 4868–4881 (2014)
26. Kayalar, S., Weinert, H.L.: Error bounds for the method of alternating projections. Math. Control Signals Syst. **1**(1), 43–59 (1988)
27. Lions, P., Mercier, B.: Splitting algorithms for the sum of two nonlinear operators. SIAM J. Numer. Anal. **16**(6), 964–979 (1979)
28. Phan, H.M.: Linear convergence of the Douglas–Rachford method for two closed sets. Optimization **65**(2), 369–385 (2016)
29. Siqueira, A.S., da Silva, R.C., Santos, L.R.: Perprof-py: A python package for performance profile of mathematical optimization software. J Open Res Softw **4**(e12), 5 (2016)
30. Svaiter, B.F.: On weak convergence of the Douglas–Rachford method. SIAM J. Control. Optim. **49**(1), 280–287 (2011)